UNCLASSIFIED

Report No. M-965 R-1

RANKING OF COMMUNITY ORGANIZATIONS

By R.M. Thrall

Assisted By R.L. Wine
and W. Feit

Directors of Project

C.H. Coombs
Associate Professor of Psychology

and

R.M. Thrall
Associate Professor of Mathematics

Project M-965 January, 1953

TABLE OF CONTENTS

RANKING OF COMMUNITY ORGANIZATIONS

1. Introduction. This paper treats some problems which came up in connection with the task of ranking social organizations on the basis of their common members. If $\mathcal{P}$ and $\mathcal{Q}$ are two organizations and if the leaders in $\mathcal{Q}$ are run of the mill members of $\mathcal{P}$ and if no high ranking members of $\mathcal{P}$ belong to $\mathcal{Q}$ we would feel intuitively that $\mathcal{P}$ is "higher" than $\mathcal{Q}$ in the social hierarchy. In addition to the relative height of two organizations we are also interested in their relative "spread". If for example, $\mathcal{P}$ and $\mathcal{Q}$ were each divided into leaders, middle group, and bottom group, and if the common members of the two organizations came from the middle group of $\mathcal{P}$ but came from all groups of $\mathcal{Q}$ we would say that $\mathcal{P}$ has greater spread than $\mathcal{Q}$.

If there were some independent measure of the height of individuals one could define height and spread relatively easily. For example, if $\mathcal{P}=\{a,\dots\}$ has r members and h(a) is a real number representing the height of a we could take

$$(1) \qquad h(\mathcal{P}) = \frac{1}{r}\sum_{a\in\mathcal{P}} h(a)$$

as a measure of the height of $\mathcal{P}$ and for the spread of $\mathcal{P}$ we could take the variance

$$(2) \qquad s(\mathcal{P}) = \frac{1}{r}\sum_{a\in\mathcal{P}} (h(a)-h(\mathcal{P}))^2$$

or the maximum difference

$$(3) \qquad s'(\mathcal{P}) = \max_{a\in\mathcal{P}} h(a) - \min_{a\in\mathcal{P}} h(a) \quad .$$

However, in general, there is no acceptable measure of the height of an individual so it is desirable to construct measures which depend only on the amounts of overlapping between subdivisions of the various organizations. Here the measures may be only relative, i.e. will merely tell which of two organiza-

tions is higher without giving a measure of the absolute height of either of the organizations being compared.

2. Relative height and spread of pairs of organizations. We suppose that each organization $\mathcal{P}$ is subdivided into n strata $\mathcal{P}_1, \ldots, \mathcal{P}_n$ starting with a highest group $\mathcal{P}_1$ and going down to a lowest group $\mathcal{P}_n$. We set

(4) $$p_i = O(\mathcal{P}_i) \;/\; O(\mathcal{P}) \qquad (i = 1, \ldots, n).$$

[For any group $\mathcal{N}$ we denote by $O(\mathcal{N})$ the number of members of $\mathcal{N}$.] Let $\mathcal{Q}$ be a second organization with strata $\mathcal{Q}_1, \ldots, \mathcal{Q}_n$ and set

(5) $$q_i = O(\mathcal{Q}_i) \;/\; O(\mathcal{Q}).$$

Now, if $\mathcal{P} \cap \mathcal{Q}$ is not empty we set

(6) $$r_{ij} = O(\mathcal{P}_i \cap \mathcal{Q}_j) \;/\; O(\mathcal{P} \cap \mathcal{Q}) \qquad (i, j = 1, \ldots, n).$$

We wish to construct some function of the p_i, q_i, r_{ij} which will tell which of $\mathcal{P}$ and $\mathcal{Q}$ is higher and to construct another function which will tell which has great spread.

A special case of importance is that of equal subdivisions, i.e.

(7) $$p_i = q_i = \sigma \qquad (i = 1, \ldots, n)$$

where $\sigma = \frac{1}{n}$. In this case the functions will depend only on the n by n matric $R = (r_{ij})$.

First consider the problem of relative height. If an individual a belongs to $\mathcal{P}_i \cap \mathcal{Q}_j$ where $i > j$, i.e., if he occupies a higher position in $\mathcal{P}$ than in $\mathcal{Q}$ then so far as this individual is concerned $\mathcal{Q}$ is higher than $\mathcal{P}$. This conclusion, of course, depends on the assumptions that the individual tries to achieve as high a position as possible in each organization to which he belongs, and that the position achieved by any individual depends only on his "height" (which we do not know). In practice neither of these assumptions is valid for

each individual although they might tend to be correct in the average.

One would say that an individual in $\mathcal{P}_1 \cap \mathcal{Q}_3$ gives more evidence of difference in height of $\mathcal{P}$ and $\mathcal{Q}$ than one in $\mathcal{P}_2 \cap \mathcal{Q}_3$. This suggests the following function

(8) $$f(R) = \sum_{i,j} (i-j)\, r_{ij}$$

and the definition $\mathcal{P}$ is higher than $\mathcal{Q}$, written $\mathcal{P} \underset{h}{>} \mathcal{Q}$ if $f(R) > 0$. If $f(R) = 0$ we say that $\mathcal{P}$ and $\mathcal{Q}$ have the same height, written $\mathcal{P} \underset{h}{=} \mathcal{Q}$.

Let $u_i = \sum_j r_{ij}$, $v_j = \sum_i r_{ij}$ $(i,j = 1,\dots,n)$. Then we have

(9) $$f(R) = \sum i(u_i - v_i).$$

To see this we write

$$f(R) = \sum_i \sum_j i\, r_{ij} - \sum_j \sum_i j\, r_{ij} = \sum_i i\, u_i - \sum_j j\, v_j = \sum i\, (u_i - v_i).$$

Next, for spread we first consider for each i the average position in $\mathcal{Q}$ of the members of $\mathcal{P}_i \cap \mathcal{Q}$. This is given by

(10) $$r_{i.} = \frac{1}{u_i} \sum_j j\, r_{ij} \qquad (i = 1,\dots,n)$$

and dually

(11) $$r_{.j} = \frac{1}{v_j} \sum_i i\, r_{ij}.$$

We then introduce the function

(12) $$g(R) = \sum_{i,j} \left[(r_{i.} - j)^2 \frac{r_{ij}}{u_i} - (r_{.j} - i)^2 \frac{r_{ij}}{v_j} \right]$$

and say that $\mathcal{P}$ has greater, equal, or less spread than $\mathcal{Q}$ according as $g(R)$ is positive, zero, or negative.

These functions $f(R)$ and $g(R)$ are not the only ones that could be used, and are introduced primarily so as to provide something concrete to work with in building a theory.

We observe that if $\mathcal{P} \cap \mathcal{Q}$ is empty (e.g. $\mathcal{P}$ = Rotary, $\mathcal{Q}$ = Lions or $\mathcal{P}$ = Methodists, $\mathcal{Q}$ = Catholics) we get no comparison. One might say if $\mathcal{P} \cap \mathcal{Q}$ is empty but $\mathcal{P} \cap \mathcal{R}$ and $\mathcal{Q} \cap \mathcal{R}$ are not empty we should somehow use the functions f and g computed first for $\mathcal{P}$ and $\mathcal{R}$ and then for $\mathcal{Q}$ and $\mathcal{R}$ and then make some comparison of $\mathcal{P}$ and $\mathcal{Q}$. Such a comparison would be justified only if the order given by f( ) is transitive, i.e. , $\mathcal{P} \underset{h}{<} \mathcal{Q}$ and $\mathcal{Q} \underset{h}{<} \mathcal{R}$ implies $\mathcal{P} \underset{h}{<} \mathcal{R}$. This is not the case as the following example shows. Take n=2 and let the only common members be those indicated in the table below.

	$\mathcal{P}$
$\mathcal{P}_1$	a_1
$\mathcal{P}_2$	a_3

	$\mathcal{Q}$
$\mathcal{Q}_1$	a_2
$\mathcal{Q}_2$	a_1

	$\mathcal{R}$
$\mathcal{R}_1$	a_3
$\mathcal{R}_2$	a_2

then clearly $\mathcal{P} \underset{h}{<} \mathcal{Q}$, $\mathcal{Q} \underset{h}{<} \mathcal{R}$, and $\mathcal{R} \underset{h}{<} \mathcal{P}$.

In spite of their limitati ns the functions f and g may be useful as building blocks in a theory.

3. Unbalanced stratification. We turn next to the case of unequal subdivisions, and consider how the numbers p_i, q_j should be introduced into the measure. The point of view we take is that, theoretically, one should always strive for equal subdivisions and the numbers p_i, q_j should be used in correcting the matrix R for any bias introduced by unequal subdivisions. We do this by constructing a new matrix $R^* = (r^*_{ij})$ whose entries are estimates, based on the observed r_{ij}, p_i q_j, of what the matrix R would have been had the subdivisions been equal. We shall describe the process of passing from R to R^* as removing the <u>bias</u> caused by use of unbalanced stratification.

We now set up some general criteria which will serve as tests for the adequacy of various bias removing constructions.

First we have some requirements for whatever functions are used to measure height and spread. If the subdivisions of $\mathcal{P}$ and $\mathcal{Q}$ are equal and the matrix R is symmetric (r.e. $r_{ij} = r_{ji}$) we require that $\mathcal{P}$ and $\mathcal{Q}$ shall have the same

height and the same spread.

Consider the case of a single organization $\mathcal{P}$ stratified by two different investigators into subsets $\mathcal{P}_1,\ldots,\mathcal{P}_n$ and $\mathcal{Q}_1,\ldots,\mathcal{Q}_n$ with corresponding proportions $p_1,\ldots,p_n$ and $q_1,\ldots,q_n$. We now apply any measures of height and spread treating $\mathcal{P}$ as though it were two organizations. We assume that the two stratifications are <u>consistent</u> in the sense that there exists a simple ordering of the individuals which is a refinement of both stratifications. This is equivalent to the requirement that for each i and j one of the two sets $\mathcal{P}_1 \cup \ldots \cup \mathcal{P}_i$, $\mathcal{Q}_1 \cup \ldots \cup \mathcal{Q}_j$ contains the other.

This assumption makes it possible to compute the r_{ij} as functions of the p_i and q_j. Clearly $r_{11} = \min(p_1,q_1)$. Proceeding inductively we get

$$(13) \qquad r_{i1} + r_{i2} + \ldots + r_{ij} = \min(p_i, \max(o, q_1 + \ldots + q_j - p_1 - \ldots - p_{i-1}))$$

and its symmetric counterpart

$$(14) \qquad r_{1j} + r_{2j} + \ldots + r_{ij} = \min(q_j, \max(o, p_1 + \ldots + p_i - q_1 - \ldots - q_{j-1})) \quad .$$

In particular

$$(15) \qquad u_i = p_i,\ v_j = q_j \quad (i,j = 1,\ldots,n).$$

Now suppose that a function $R^* = b\,(R, p_1,\ldots,p_n, q_1,\ldots,q_n)$ is proposed as a bias removing construction. Complete removal of bias for two consistent stratifications of a single organization $\mathcal{P}$ would lead to

$$(16) \qquad R^* = \sigma\, I_n,$$

since this is what would be obtained from consistent equal subdivisions. However, if this were not achieved one might ask that R^* be symmetric, i.e.

$$(17) \qquad R^* = (R^*)^{Tr} \quad ,$$

here Tr indicates transposed matrix.

If R^* is symmetric then we at least are assured that we will not be claiming that an organization is higher than (or has more spread than) itself.

Finally with reference to particular measures $f(R)$ and $g(R)$ of relative height and of relative spread we might ask that

(18) $$f(R^*) = 0 \text{ and } g(R^*) = 0.$$

Note that (16) guarantees (17) and (18), and (17) guarantees (18) whereas knowing that (18) is true for one pair $f(\ )$ and $g(\ )$ gives no guarantee that it will hold for other measures. Thus it is highly desirable to achieve (16) and (17).

4. Bias removal by matrix multiplication. If $\mathcal{P}$ and $\mathcal{Q}$ are two stratified organizations the bias in the comparison matrix R can be regarded as coming from unbalance in both stratifications. It is natural to ask if we can remove the bias in two steps, one to care for the unbalance in the $\mathcal{P}_i$ and one for the unbalance in the $\mathcal{Q}_j$; and moreover, so that the first step is independent of $\mathcal{Q}$ and the second step is independent of $\mathcal{P}$. In other words we ask if it is possible to associate a correction operation with each stratification of each organization.

Let $\mathcal{P}_1, \ldots, \mathcal{P}_n$ be a stratification of $\mathcal{P}$ and let $\Sigma 1, \ldots, \Sigma n$ be a consistent equal stratification. [Here we assume either that $O(\mathcal{P})$ is divisible by n or that $O(\mathcal{P})$ is large enough in comparison with n so that approximately equal subdivisions are possible. For example, the case $O(\mathcal{P}) = 10$ and $n = 6$ would be ruled out, but $O(\mathcal{P}) = 50$ and $n = 3$ would be accepted. Actually the corrections obtained can be applied in every case but the justification depends on the existence of the Σi.]

Now suppose that p_{ij} is the proportion of $\mathcal{P}_j$ which lies in Σi, i.e. $p_{ij} = O(\mathcal{P}_j \cap \Sigma i) / O(\mathcal{P}_j)$. Clearly $\mathcal{P}_j = \bigcup_i (\mathcal{P}_j \cap \Sigma i)$, hence

(19) $$\sum_i p_{ij} = 1.$$

Since $p_j = O(\mathcal{P}_j) / O(\mathcal{P})$ and $\sigma = O(\Sigma i) / O(\mathcal{P})$ we have from $\Sigma i = \bigcup_j (\mathcal{P}_j \cap \Sigma i)$ that

$$\frac{O(\Sigma i)}{O(\mathcal{P})} = \frac{1}{O(\mathcal{P})} \sum_j O(\mathcal{P}_j \cap \Sigma i) = \sum_j \frac{O(\mathcal{P}_j \cap \Sigma i)}{O(\mathcal{P}_j)} \cdot \frac{O(\mathcal{P}_j)}{O(\mathcal{P})}$$

or

(20) $$\sigma = \sum_j p_{ij} p_j \quad .$$

Because of our assumption of consistency between the $\mathcal{P}_j$ and Σi stratifications we get

(21) $$p_{ij}\, p_j = \begin{cases} 0 & \text{if } p_1 + \dots + p_j \le (i-1)\sigma \\ p_1 + \dots + p_j - (i-1)\sigma & \text{if } p_1 + \dots + p_{j-1} \le (i-1)\sigma < p_1 + \dots + p_j \le i\sigma \\ \sigma & \text{if } p_1 + \dots + p_{j-1} \le (i-1)\sigma \text{ and } i\sigma \le p_1 + \dots + p_j \\ p_j & \text{if } (i-1)\sigma \le p_1 + \dots + p_{j-1} \text{ and } p_1 + \dots + p_j \le i\sigma \\ i\sigma - (p_1 + \dots + p_{j-1}) & \text{if } (i-1)\sigma \le p_1 + \dots + p_{j-1} \le i\sigma \le p_1 + \dots + p_j \\ 0 & \text{if } i\sigma \le p_1 + \dots + p_{j-1} \end{cases}$$

We define the bias correction for $\mathcal{P}$ to be the replacement of the matrix R by the matrix R' where

(22) $$r'_{ij} = \sum p_{i\nu}\, r_{\nu j} .$$

This has the effect of splitting $r_{\nu j}$ into the same proportions as $\mathcal{P}_\nu$ is split by the Σ_j.

In matrix form (22) becomes

(23) $$R' = P\, R$$

where $$P = (p_{ij}) .$$

To remove the bias caused by inequalities in the stratification of $\mathcal{Q}$ we form the matrix $Q = (q_{ij})$ where $q_{ij} = O(\mathcal{Q}_j \cap \Sigma i) / O(\mathcal{Q}_j)$ and then replace R' by R* where

(24) $$r^*_{ij} = \sum r'_{i\nu}\, q_{j\nu} \quad .$$

In matrix form

(25) $$R^* = R'Q^{Tr} = PRQ^{Tr}$$

The associativity of matrix multiplication, i.e. $(PR)\,Q^{Tr} = P\,(RQ^{Tr})$, shows that the final result is independent of the order of the corrections.

The following example illustrates the procedure. Let $\mathcal{P} = \mathcal{Q}$ have two consistent stratifications in which $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$, and $q_1 = \frac{1}{6}$, $q_2 = \frac{1}{2}$, $q_3 = \frac{1}{3}$ then (see(13) and(21)for computations of R, P, and Q)

(26) $$R = \left\| \begin{matrix} \frac{1}{6} & \frac{1}{3} & 0 \\ 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{1}{6} \end{matrix} \right\|, \quad P = \left\| \begin{matrix} \frac{2}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \end{matrix} \right\|, \quad Q = \left\| \begin{matrix} 1 & \frac{1}{3} & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & 1 \end{matrix} \right\|$$

and

(27) $$R^* = P\,R\,Q^{Tr} = \left\| \begin{matrix} \frac{5}{27} & \frac{4}{27} & 0 \\ \frac{13}{108} & \frac{7}{54} & \frac{1}{12} \\ \frac{1}{36} & \frac{1}{18} & \frac{1}{4} \end{matrix} \right\| .$$

This result is no accident. For let $\mathcal{P}_1$ and $\mathcal{Q}_1$ be any two consistent stratification of an organization and let $R^* = PRQ^{Tr}$. Then $u_i^* = v_j^* = \sigma$ where $u_i^* = \sum_{\nu} r_{i\nu}$, etc. We have

(28) $$u_i^* = \sum_{\nu} r_{i\nu}^* = \sum_{\nu,\mu,\lambda} p_{i\mu} r_{\mu\lambda} q_{\nu\lambda}.$$

Now by (19), $\sum_{\nu} q_{\nu\lambda} = 1$, hence $u_i^* = \sum_{\mu,\lambda} p_{i\mu} r_{\mu\lambda}$.
By (15) $\sum_{\lambda} r_{\mu\lambda} = p_{\mu}$, hence $u_i^* = \sum_{\mu} p_{i\mu} p_{\mu}$ and by (20) this is σ. The proof for $v_j^* = \sigma$ is similar.

It now follows from (9) that $f(R^*) = 0$, i.e. condition (18) holds for the $f(\ )$ defined in (8). However, for R^* given by (27) we do not have $g(R^*) = 0$ hence the second half of (18) fails for the $g(\ )$ of (12). Of course (27) shows that neither (17) nor (16) hold for this type of bias correction.

More generally we ask for ways in which a matrix P can be assigned to each $(p_1, \ldots, p_n)$ so that wherever $\mathcal{P}_1, \ldots, \mathcal{P}_n$ and $\mathcal{Q}_1, \ldots, \mathcal{Q}_n$ are two consistent stratifications of an organization, that $R^* = PRQ^{Tr}$ has $u_i^* = v_j^* = \sigma$ $(i,j = 1, \ldots, n)$. In view of (28) it is sufficient to have equations (19) and (20) for each P.

We naturally require that $P = I_n$ if $p_1 = \ldots = p_n = \sigma$ since then there is no bias to be corrected. Now if $q_1 = \ldots = q_n = \sigma$, i.e. $Q = I_n$ we get
$u_i^* = \sum_\nu r_{i\nu}^* = \sum_{\mu,\nu} p_{i\mu} r_{\mu\nu} = \sum_\mu p_{i\mu} p_\mu$ since $\sum_\nu r_{i\nu} = p_i$.
This shows that condition (20) is also necessary. We have not been able to determine if condition (19) is necessary.

5. Bias removal by simultaneous corrections. Although this first method is not entirely satisfactory it points the way to a second imporved method. One objection to this first method is illustrated by our example above. Here $p_1 = \frac{1}{2}$ was too large and we corrected each r_{ij} to take account of this. Now, since $\Sigma 1 \subseteq \mathcal{P}_1$ and $\Sigma 1$ is higher than $\Sigma 2$ perhaps we should have assumed that all members of $\mathcal{P}_1 \cap \mathcal{Q}_1$ belonged to $\Sigma 1$ and have made any necessary corrections on the later $r_{1\nu}$. The following method incorporates this idea.

Suppose that $\mathcal{P}_1 \cup \ldots \cup \mathcal{P}_{h-1} \subset \Sigma 1 \cup \ldots \cup \Sigma i \subseteq \mathcal{P}_1 \cup \ldots \cup \mathcal{P}_h$, i.e. $p_1 + \ldots + p_{h-1} < i\sigma \leq p_1 + \ldots + p_h$. It seems reasonable to require that

$$u_1^* + \ldots + u_i^* = u_1 + \ldots + u_{h-1} + \alpha u_h \tag{29}$$

where $\alpha = (i\sigma - (p_1 + \ldots + p_{h-1})) / p_h$. This is equivalent to assuming that for each organization $\mathcal{Q}$ we have

$$\frac{o((\Sigma 1 \cup \ldots \cup \Sigma i) \cap \mathcal{P}_h \cap \mathcal{Q})}{o(\mathcal{P}_h \cap \mathcal{Q})} = \frac{o((\Sigma 1 \cup \ldots \cup \Sigma i) \cap \mathcal{P}_h)}{o(\mathcal{P}_h)} . \tag{30}$$

We introduce a more detailed notation to care for all i. Let h_i be defined by the equation

(31) $$p_1 + \dots + p_{h_i-1} < i\sigma \leq p_1 + \dots + p_{h_i} \quad (i=1, \dots, n)$$

and let

(32) $$\alpha_i = (i\sigma - p_1 - \dots - p_{h_i-1}) / p_{h_i} \quad (i=1, \dots, n).$$

Similarly suppose that k_j and β_j are defined by

(33) $$q_1 + \dots + q_{k_j-1} < j\sigma \leq q_1 + \dots + q_{h_j} \quad (j=1, \dots, n)$$

and

(34) $$\beta_j = (j\sigma - q_1 - \dots - q_{h_j-1}) / q_{k_j} \quad (j=1, \dots, n).$$

We now define u_i^* and v_j^* inductively by the equations

(35) $$u_i^* = u_1 + \dots + u_{h_i-1} + \alpha_i u_{h_i} - u_1^* - \dots - u_{i-1}^* \quad (i=1, \dots, n)$$

and

(36) $$v_j^* = v_1 + \dots + v_{k_j-1} + \beta_j v_{h_j} - v_1^* - \dots - v_{j-1}^* \quad (j=1, \dots, n).$$

Our next task is to define a matrix $R = \| r_{ij}^* \|$ for which the u_i^* and v_j^* are respectively the row and column sums. Our definitions are inductive; to determine r_{ij}^* we assume that all r_{hj}^*, r_{ik}^* with $h<i$ or $k<j$ are already known.

First, we set

(37) $$c_{ij} = u_{h_i} - r_{h_i 1} - \dots - r_{h_i k_j} \quad (i,j=1, \dots, n)$$

(38) $$d_{ij} = v_{k_j} - r_{1k_j} - \dots - r_{h_i k_j} \quad (i,j=1, \dots, n)$$

(39) $$c_{ij}^* = u_i^* - r_{i1}^* - \dots - r_{ij-1}^* \quad (i,j=1, \dots, n)$$

(40) $$d_{ij}^* = v_j^* - r_{1j}^* - \dots - r_{i-1j}^* \quad (i,j=1, \dots, n)$$ and then

(41) $$r_{ij}^* = \min \{ c_{ij}^*, d_{ij}^*, \max(c_{ij}^* - \alpha_1 c_{ij}, d_{ij}^* - \beta_j d_{ij}) \} .$$

Note, that for $i = j = 1$ (41) reduces to

(42) $$r^*_{11} = \min\left\{u^*_1,\ v^*_1,\ \max(u^*_1 - \alpha_1 c_{11},\ v^*_1 - \beta_1 d_{11})\right\}$$

which gives a basis for the inductive definition.

We now apply this second method of correction to the test case of two consistent subdivisions of an organization $\mathcal{P}$. We now have from (15) $u_i = p_i$, $v_j = q_j$ and hence from (32) and 34) we get

(43) $$u^*_i = v^*_j = \sigma \quad (i,j = 1, \ldots, n).$$

Next, from (13) and (14) we get

$$c_{ij} = p_{h_i} - \min(p_{h_i}, q_1 + \cdots + q_{k_j} - p_1 - \cdots - p_{h_i - 1})$$

and

$$d_{ij} = q_{k_j} - \min(q_{k_j}, p_1 + \cdots + p_{h_i} - q_1 - \cdots - q_{k_j - 1}).$$

Thus $c_{ij} = 0$ if $p_1 + \cdots + p_{h_i} \leq q_1 + \cdots + q_{k_j}$ and otherwise $d_{ij} = 0$; hence $\max(c^*_{ij} - \alpha_i c_{ij},\ d^*_{ij} - \beta_j d_{ij}) \geq \min(c^*_{ij},\ d^*_{ij})$ from which it follows that

(44) $$r^*_{ij} = \min(c^*_{ij},\ d^*_{ij}) \quad (i,j = 1, \ldots, n).$$

In particular $r_{11} = \sigma$. We now take as an induction hypothesis that

(45) $$r^*_{\nu\mu} = \sigma\,\delta_{\nu\mu}$$

for all $(\nu, \mu) \neq (i,j)$ such that $\nu \leq i$, $\mu \leq j$. Then

(46) $$c_{ij} = \sigma - \sum_{m=1}^{j-1} \sigma\,\delta_{i\mu} = \begin{cases} \sigma & \text{if } j \leq i \\ 0 & \text{if } j > i \end{cases}$$

and

(47) $$d^*_{ij} = \begin{cases} \sigma & \text{if } i \leq j \\ 0 & \text{if } i > j \end{cases}$$

Hence

$$r^*_{ij} = \min(c^*_{ij},\ d^*_{ij}) = \sigma\,\delta_{ij}.$$

This completes the induction argument and establishes the equality $R^* = \sigma I_n$.

Thus we see that this second method of removing bias meets our strongest test condition (16), whereas the first method gives only the weaker condition (18) and this only for the f(R) given by (9).

6. **Some relations between the unbiased matrices and their functions.** Let R_1^* be the comparison matrix after bias is removed by matrix multiplication. Let R_2^* be the comparison matrix after bias is removed by simultaneous corrections. Let $\bar{u}_i^*$ and $\bar{v}_j^*$ denote the row and column sums respectively for R_2^* (u_i^* and v_j^* will refer to R_1^*).

First we will show that

$$u_i^* = \bar{u}_i^*. \tag{48}$$

To prove this it is convenient to have (21) written in a different form. After dividing both members of (21) by p_j and introducing the h_i defined in (31) we get

$$p_{ij} = \begin{cases} 0 & \text{if } j < h_{i-1} \\ 1-\alpha_{i-1} & \text{if } j = h_{i-1} < h_i \\ \sigma/p_j & \text{if } h_{i-1} = j\ h_i \\ 1 & \text{if } h_{i-1} < j < h_i \\ \alpha_i & \text{if } h_{i-1} < j = h_i \\ 0 & \text{if } h_i < j \end{cases},$$

where $\alpha_i = \dfrac{i\sigma - p_1 - \cdots - p_{h_i-1}}{p_{h_i-1}}$

and

$$1-\alpha_{i-1} = 1 - \frac{(i-1)\sigma - p_1 - \cdots - p_{h_{i-1}-1}}{p_{h_{i-1}}} = \frac{p_1 + \cdots + p_{h_{i-1}} - (i-1)\sigma}{p_{h_{i-1}}}.$$

Remembering that $R_1^* = PRQ^{Tr}$, and using (28) and (19), we write

$$u_i^* = \sum_{\nu,\mu,\lambda} p_{i\nu}\, r_{\nu\mu}\, q_{\lambda\mu} = \sum_{\nu,\mu} p_{i\nu}\, r_{\nu\mu} = \sum_{\nu} p_{i\nu}\, u_\nu . \tag{50}$$

We have from (35), using the new notation,

(51) $$\bar{u}_i^* = u_1 + \dots + u_{h_i - 1} + \alpha_i u_{h_i} - (\bar{u}_1^* + \dots + \bar{u}_{i-2}^* + \bar{u}_{i-1}^*)$$

and

(52) $$\bar{u}_{i-1}^* = u_1 + \dots + u_{h_{i-1}-1} + \alpha_{i-1} u_{h_{i-1}} - (\bar{u}_1^* + \dots + \bar{u}_{i-2}^*).$$

Hence, on substituting (52) in (51), we write

(53) $$\bar{u}_i^* = (1 - \alpha_{i-1}) u_{h_{i-1}} + u_{h_{i-1}+1} + \dots + u_{h_i - 1} + \alpha_i u_{h_i}, \text{ if } h_{i-1} < h_i.$$

if $h_{i-1} = h_i$, then

(54) $$\bar{u}_i^* = (\alpha_i - \alpha_{i-1}) u_{h_i} = \frac{\sigma}{p_{h_i}} \cdot u_{h_i} = p_{ih_i} u_{h_i},$$

since

$$\alpha_i - \alpha_{i-1} = \frac{i\sigma - (p_1 + \dots + p_{h_i - 1})}{p_{h_i}} - \frac{(i-1)\sigma - (p_1 + \dots + p_{h_i - 1})}{p_{h_i}} = \frac{\sigma}{p_{h_i}}.$$

So in any case, we see by (49) that

(55) $$\bar{u}_i^* = \sum_j p_{ij} u_j$$

for if $j < h_{i-1}$ or $p > h_i$ then $p_{ij} = 0$. Thus $u_i^* = \bar{u}_i^*$. By a similar argument

(56) $$v_j^* = \bar{v}_j^*.$$

Next we will prove that

(57) $$f(R_1^*) = f(R_2^*).$$

By (9), (48) and (56)

$$f(R_1^*) = \sum i(u_i^* - v_i^*) = \sum i(\bar{u}_i^* - \bar{v}_i^*) = f(R_2^*).$$

7. Some illustrative examples. The particular cases of equal consistent and unbalanced consistent stratifications have been covered in the development of the theory. We will now illustrate the theory for unbalanced stratifications when a comparison matrix R is given.

Example 1. Let $\mathcal{P}$ and $\mathcal{Q}$ be two organizations with two stratifications each where $p_1 = \frac{2}{5}$, $p_2 = \frac{3}{5}$, $q_1 = \frac{4}{7}$, $q_2 = \frac{3}{7}$ and

$$R = \begin{Vmatrix} \frac{2}{10} & \frac{4}{10} \\ \frac{3}{10} & \frac{1}{10} \end{Vmatrix} = \frac{1}{10} \begin{Vmatrix} 2 & 4 \\ 3 & 1 \end{Vmatrix}.$$

For example, $\mathcal{P}$ may have 50 members with 20 belonging to $\mathcal{P}_1$ and 30 to $\mathcal{P}_2$, and $\mathcal{Q}$ may contain 70 members with 40 belonging to $\mathcal{Q}_1$ and 30 belonging to $\mathcal{Q}_2$. Of the 10 members $\mathcal{P}$ and $\mathcal{Q}$ have in common 2 of the members in $\mathcal{P}_1$ are also in $\mathcal{Q}_1$, 4 of the members in $\mathcal{P}_1$ are in $\mathcal{Q}_2$, 3 members in $\mathcal{P}_2$ are in $\mathcal{Q}_1$, and 1 member in $\mathcal{P}_1$ is in $\mathcal{Q}_2$.

Now $u_1 = \frac{6}{10}$, $u_2 = \frac{4}{10}$, $v_1 = \frac{5}{10}$ and $v_2 = \frac{5}{10}$. So from (9),

$$f(R) = 1\left(\frac{6}{10} - \frac{5}{10}\right) + 2\left(\frac{4}{10} - \frac{5}{10}\right) = -\frac{1}{10}.$$

Thus we conclude that $\mathcal{Q}$ is higher than $\mathcal{P}$ for the biased matrix R.

Further, $r_{1.} = \frac{5}{3}$, $r_{2.} = \frac{5}{4}$, $r_{.1} = \frac{8}{5}$ and $r_{.2} = \frac{6}{5}$. Thus by (12)

$$g(R) = \left(\frac{5}{3} - 1\right)^2 \frac{2}{6} + \left(\frac{5}{3} - 2\right)^2 \frac{4}{6} + \left(\frac{5}{4} - 1\right)^2 \frac{3}{4} + \left(\frac{5}{4} - 2\right)^2 \frac{1}{4}$$
$$-\left(\frac{8}{5} - 1\right)^2 \frac{2}{5} - \left(\frac{8}{5} - 2\right)^2 \frac{3}{5} - \left(\frac{6}{5} - 1\right)^2 \frac{4}{5} - \left(\frac{6}{5} - 2\right)^2 \frac{1}{5} = \frac{7}{720}.$$

So we say $\mathcal{P}$ has greater spread than $\mathcal{Q}$.

We wish also to find the unbiased matrix R_1^*. From the definitions of p_{ij} and q_{ij}, using the particular illustration with 50 and 70 members respectively, we find that

$$P = \begin{Vmatrix} 1 & \frac{1}{6} \\ 0 & \frac{5}{6} \end{Vmatrix} \quad \text{and} \quad Q = \begin{Vmatrix} \frac{7}{8} & 0 \\ \frac{1}{8} & 0 \end{Vmatrix}.$$

[In general, P and Q may also be found by (21).] So

$$R_1^* = (PR)\,Q^{Tr} = \frac{1}{60} \begin{Vmatrix} 15 & 25 \\ 15 & 5 \end{Vmatrix} \cdot \begin{Vmatrix} \frac{7}{8} & \frac{1}{8} \\ 0 & 1 \end{Vmatrix} = \frac{1}{96} \begin{Vmatrix} 21 & 43 \\ 21 & 11 \end{Vmatrix}.$$

Hence by (9) and (12),

$$f(R_1^*) = -\frac{11}{48} \sim -.229$$

and

$g(R_1^*) \sim .034$, where $\sim$ is an approximation symbol.

Finally, R_2^* is found in the following way. Since $\alpha = \frac{1}{2}$, for $i = 1$, $\frac{2}{5} < 1 \cdot \frac{1}{2} \leq \frac{2}{5} + \frac{3}{5}$ and thus $h_1 = 2$. So by (32) $\alpha_1 = \dfrac{\frac{1}{2} - \frac{2}{5}}{\frac{3}{5}} = \frac{1}{6}$.

Similarly when $i = 2$, $h_2 = 2$ and $\alpha_2 = 1$, when $j = 1$, $k_1 = 1$ and $\beta_1 = \frac{7}{8}$ and when $j = 2$, $k_2 = 2$ and $\beta_2 = 1$. Thus by (35) and (36), $u_1^* = u_1 + \alpha_1 u_2 = \frac{2}{3}$, $u_2^* = \frac{1}{3}$, $v_1^* = \frac{7}{16}$ and $v_2^* = \frac{9}{16}$. Further using (37) and (38), $c_{11} = u_2 = r_{21} = \frac{1}{10}$, $c_{12} = 0$, $c_{21} = \frac{1}{10}$, $c_{22} = 0$, $d_{11} = 0 = d_{12} = d_{21} = d_{22}$. From (39), (40), and (41) we get in order

$$c_{11}^* = u_1^* = \frac{2}{3},$$

$$d_{11}^* = v_1^* = \frac{7}{16},$$

$$r_{11}^* = \min\left\{\frac{2}{3}, \frac{7}{16}, \max\left[\frac{2}{3} - \frac{1}{6} \cdot \frac{1}{10}, \frac{7}{16}\right]\right\} = \min\left\{\frac{2}{3}, \frac{7}{16}, \frac{39}{60}\right\} = \frac{7}{16},$$

$$c_{12}^* = u_1^* - r_{11}^* = \frac{11}{48}$$

$$d_{12}^* = v_2^* = \frac{9}{16},$$

$$r_{12}^* = \min\left\{\frac{11}{48}, \frac{9}{16}, \max\left[\frac{11}{48} - 0, \frac{9}{16}\right]\right\} = \frac{11}{48},$$

etc.

so that

$$R_2^* = \frac{1}{48} \begin{Vmatrix} 21 & 11 \\ 0 & 16 \end{Vmatrix} .$$

Hence

$$g(R_2^*) \sim - .016.$$

In this example the two bias corrections lead us to oppostie conclusions for relative spread.

Example 2. Let $\mathcal{P}$ and $\mathcal{Q}$ be two organizations with three stratifications each where $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$, $q_1 = \frac{1}{6}$, $q_2 = \frac{1}{2}$, $q_3 = \frac{1}{3}$ and

$$R = \frac{1}{10} \begin{Vmatrix} 2 & 4 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{Vmatrix} .$$

By the methods used in example 1 we find that

$$R_1^* = \frac{1}{180} \begin{Vmatrix} 40 & 32 & 0 \\ 23 & 22 & 0 \\ 21 & 6 & 36 \end{Vmatrix}$$

and

$$R_2^* = \frac{1}{60} \begin{Vmatrix} 24 & 0 & 0 \\ 4 & 11 & 0 \\ 0 & 9 & 12 \end{Vmatrix} .$$

We thus find the relative height and relative spread numbers as given in Table1.

Table 1

	g	f
R	.062	-.200
R_1^*	.225	$\frac{13}{60} \sim .217$
R_2^*	.071	

For relative spread all three tests give the same result, i.e. $\mathcal{P}$ has greater spread than $\mathcal{Q}$. As for relative height, the unbiased matrices give a result which differs from that of the biased matrix.

Example 3. Let $\mathcal{P}$ and $\mathcal{Q}$ be two organizations with three stratifications each where $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$, $q_1 = \frac{1}{6}$, $q_2 = \frac{1}{2}$, $q_3 = \frac{1}{3}$, and

$$R = \frac{1}{2} \begin{Vmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{Vmatrix} .$$

We find

$$R_1^* = \frac{1}{6} \begin{Vmatrix} 0 & 0 & 2 \\ 0 & 0 & 1 \\ 3 & 0 & 0 \end{Vmatrix}$$

and

$$R_2^* = \frac{1}{6} \begin{Vmatrix} 2 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 3 \end{Vmatrix} .$$

From those values of R we get Table 2.

Table 2

	g	f
R	0	0
R_1^*	$-\frac{2}{9}$	$\frac{1}{6}$
R_2^*	$-\frac{2}{9}$	

Without bias removing matrices we would conclude from Table 2 that the two organizations have the same relative height and the same relative spread. But the unbiased matrices show that $\mathcal{P}$ is higher than $\mathcal{Q}$ whereas $\mathcal{Q}$ has greater spread than $\mathcal{P}$.

Example 4. Let $\mathcal{P}$ and $\mathcal{Q}$ be two organizations with four stratifications each where

$$p_1 = \frac{1}{3},\ p_2 = \frac{1}{4},\ p_3 = \frac{1}{6},\ p_4 = \frac{1}{4},$$

$$q_1 = \frac{1}{3},\ q_2 = \frac{1}{6},\ q_3 = \frac{1}{6},\ q_4 = \frac{1}{3},$$

and

$$R = \frac{1}{10}\begin{Vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 2 \end{Vmatrix} .$$

Thus

$$R_1^* = \frac{1}{480}\begin{Vmatrix} 27 & 9 & 9 & 27 \\ 9 & 35 & 43 & 33 \\ 0 & 16 & 20 & 12 \\ 108 & 36 & 24 & 72 \end{Vmatrix}$$

and

$$R_2^* = \frac{1}{120}\begin{Vmatrix} 9 & 9 & 0 & 0 \\ 6 & 8 & 16 & 0 \\ 12 & 0 & 0 & 0 \\ 9 & 7 & 8 & 36 \end{Vmatrix}$$

along with R give the numbers found in Table 3.

Table 3

	g	f
R	1.528	.300
R_1^*	-.043	.450
R_2^*	-.965	

From Table 3 we conclude that $\mathcal{P}$ is higher than $\mathcal{Q}$ in all cases. The two bias removing tests indicate that $\mathcal{Q}$ has greater spread than $\mathcal{P}$, but the given matrix indicates that $\mathcal{P}$ has greater spread than $\mathcal{Q}$.

These examples indicate the importance of the bias removing processes and also that each method for removing bias has value. In example 1 we see that bias removed by simultaneous corrections is necessary for relative spread. In examples 2, 3, and 4 we see that some sort of bias removing method is needed. If we wish only to test relative height, then the R_2^* matrix is unnecessary. Further in all the examples $g(R_2^*) \leq g(R_1^*)$. Whether this is generally true is still an open question.